WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

EXPERIENCE WITH A MODEL OF SEQUENTIAL DIAGNOSIS

296-67

G. Anthony Gorry*

and

G. Octo Barnett**

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

EXPERIENCE WITH A MODEL OF SEQUENTIAL DIAGNOSIS

296-67

G. Anthony Gorry*

and

G. Octo Barnett**

*Assistant Professor, Sloan School of Management, M.I.T.

**Director, Laboratory of Computer Science, Massachusetts General Hospital

Work reported herein was supported (in part) by NIH grant GM-15287-03 and (in part) by Project MAC, an M.I.T. research program sponsored by the Advanced Research Projects Agency, Department of Defense, under Office of Naval Research Contract Number Nonr-4102(01).

In recent years, a number of studies of the use of computer programs in diagnosis have been performed. Central to each of these efforts has been the development of an explicit, precisely formulated procedure for diagnosis. Such a development is a prerequisite for computer programs of this type. In general, attention has been focused on models of the inference function of diagnosis, the development of a diagnosis from the given set of clinical signs. Some interesting probabilistic models have been developed which employ Bayes rule.

Bayes rule has understandable appeal for use in such a model. First, it permits the use of probabilities in inference. This is preferable to a deterministic approach, because it reflects some of the basic uncertainties of diagnosis. Also, Bayes rule provides a rational means for considering both a priori belief about the incidence of various diseases and the evidence embodied in the clinical signs in a given case. Finally, the formulation of the inference function in terms of Bayes rule is particularly suited for incorporation into a computer program. Given the necessary statistical data, the problem of inference is thereby reduced to a problem of computation.

While the Bayesian model is well suited for computer diagnosis, there are certain problems associated with its use. First, the model requires that extensive statistical data be available for the given area. The

collection and processing of this data may be a very formidable task. There are also problems in properly accounting for the dependence of various signs and the possibility of the simultaneous occurrence of more than one disease. In spite of these difficulties, a number of investigations of the use of the Bayesian model have obtained encouraging results.

Since Ledley and Lusted first discussed the potential value of probabilistic models of diagnostic inference in 1959 (1), the Bayesian model has been applied to several distinct problems of diagnosis. Warner and his associates have conducted extensive studies of this model in the diagnosis of congenital heart disease (2, 3); Overall and Williams have explored its use in the diagnosis of thyroid condition (4); and Lodwick has investigated its applicability to the diagnosis of bone tumors (5). In each case, these workers reported that a program based on this model performed at a level approaching that of an expert in the respective field.

These studies all employed the same basic approach. A given set of tests are performed on the patient, and the test results become the input to the program. Through the use of Bayes rule, the program computes the conditional probability distribution for the diseases in question. This distribution constitutes the diagnosis provided by the program.

This approach, however, reflects only one aspect of diagnosis, that of inference. There is another important task in diagnosis -- the determination of an appropriate sequence of diagnostic tests to perform on the patient. In some cases, this may simply involve selecting a 'good'

Figure 1

The Use of Bayes Rule for Inference

Bayes rule can be employed to obtain the conditional probability for a disease given the observation of a new attribute as follows:

$$(1) \qquad P''(D_k/E'') = \frac{P(S_i/D_k, E')\, P'(D_k/E')}{\sum_j P(S_i/D_j, E')\, P'(D_j/E')}$$

where $P'(D_k/E')$ is the probability of D_k given the total experience to date, E', but before the observation of the attribute S_i.

$P(S_i/D_k, E')$ is the conditional probability of attribute S_i given D_k and E'. Notice that if this probability does not depend on other observations during the diagnosis, it can be replaced by $P(S_i/D_k)$.

and $P''(D_k/E'')$ is the probability for D_k given the new, increased experience E''. This new experience includes E' and the observation S_i.

Several comments about this mode of inference are relevant here. Before any observations have been made in a diagnosis, $P(D_k/E')$ is the a priori probability of D_k. This probability is continually updated as knowledge is gained about the specific case. Thus E' reflects both observations on the given patient, and experience with this set of diseases in general.

The updated probability, $P''(D_k/E'')$, becomes $P'(D_k/E')$ for the next stage of the inference process. That is, the latest view of the problem as embodied in the distribution over the diseases is used in the analysis of a new observation.

series of questions to ask, while in other cases, it may mean the selection of one or several from a set of elaborate and costly laboratory tests. In what follows, we will be concerned with the second aspect of diagnosis. In this discussion, we will employ some terms which require definition. First, an attribute is a sign or symptom which can provide information for the diagnosis. An attribute is binary-valued (it is either present or absent). A test is the means employed to detect the presence or absence of one or more attributes. For example, the test for the attribute "patient between 10 and twenty years of age" might be a simple question. The test for some other attribute may be a laboratory procedure. Those aspects of diagnosis which are concerned with the selection of a test or sequence of tests will be referred to collectively as the test selection function.

Diagnosis, then, consists of two major functions, inference and test selection. The computer programs mentioned above, however, deal only with the first of these functions. As a result, the basic structure of these programs is inadequate for dealing with broader problems of diagnosis. A more complete model of diagnosis must provide for the interaction of these two functions. A physician seldom has sufficient information initially to make a satisfactory diagnosis. He can use the information at hand, however, to form a tentative view of the problem. Using this current view of the problem in conjunction with his medical knowledge and experience, he can select a testing strategy which he expects to yield significant information. Since he often performs these tests sequentially, he has the opportunity to modify his intended testing strategy in the light of new and perhaps unexpected test results. A model of diagnosis should reflect these possibilities.

In considering a particular test, the physician should weigh the expected value of the test results against the expected cost of the test. Because tests can be costly (in terms of patient discomfort, time of skilled persons, money, etc.) diagnostic tests should be kept to a minimum. On the other hand, the physician seeks to minimize the consequences of possible misdiagnoses. In general, the probability of such an error is reduced as more information about the patient is available. Hence the physician may wish to perform a large number of tests to reduce his uncertainty about the condition of the patient. Because these are two contradictory objectives, he must strike a balance between the two. In this view, then, an appropriate model of diagnosis is one which reflects this sequential decision problem confronting the physician.

B. A Program for Sequential Diagnosis

In this section, a model for sequential diagnosis is presented. The model is intended to serve as the basis for a computer program for sequential diagnosis. Such a program has been implemented as part of a system for computer-aided diagnosis and is discussed in detail elsewhere. (6, 7) For convenience, we will discuss this model in terms of a computer realization of it. Note that the model is intended for use on a computer, and it is not intended as a descriptive model of the decision-making processes of physicians. The strategies employed in the model were chosen with regard for the capabilities and limitations of a digital computer.

There are three parts to this section. In the first part, the requirements for the information base for the program are described. It is this information base which constitutes the medical "experience" of the program. The second section is devoted to a description of the inference function of the program. The inference performed by the program is probabilistic and is based on the Bayesian model. In the third section, the test selection function of the program is explained.

Our concern here is not with the details of the realization of a program of the type described. These details are available in the references cited above. Instead we will attempt to indicate an overall approach to the problem of computer diagnosis. It is important to note a significant difference between the program described here and those mentioned above. This program operates in an interactive mode. That is, the user of the program engages in a dialogue with the program. Although the program selects tests to be performed on the patient, it is the user of the program, the physician, who performs these tests, and reports the results to the program. Through this dialogue, the program can continually advise the physician as it improves its current view of the problem.

1. Information Requirements for the Program

The basic mode of inference employed by the program is probabilistic, and a major portion of the information requirement consists of probabilities. Like the programs mentioned above, the program employs a Bayesian analysis of attributes as its central inference mechanism. Therefore, the program

requires access to the a priori probabilities of the diseases being considered as well as the conditional probabilities for each of the relevant attributes given each of the diseases. The program also must be provided with a set of tests with which it can obtain information about a particular diagnostic problem. In the model, the program knows the name of each test, the attributes which can be detected by the test, and the cost of the test. In this model, there must be at least one test which can detect the presence or absence[1] of each attribute which is considered relevant to diagnosis. There may be, however, several tests which can reveal the same attribute. Tests may range from such simple ones as asking the patient's age to complex laboratory procedures. In this model, it is assumed that no errors are made in performing tests. Of course, such may not be the case, but in any event, the inference function makes no allowance for possible error. Uncertainty about test results or unreliable tests can be accommodated by the underlying model, but the manner in which this can be done will not be discussed here.

In addition to the information outlined above, the program also requires an indication of the relative seriousness of possible misdiagnoses. In the model, this information takes a particular form. For each ordered pair of diseases, the program is given a number which represents the cost of misdiagnosing one disease as the other. For example, the cost of misdiagnosing a malignant tumor as benign might be 1,000,000 while cost of the opposite misdiagnosis might be 10,000. It is assumed that these costs are in the same units as the costs of tests. For the program, the total of this information constitutes the medical experience upon which it can draw

---

[1] Note that the absence of an attribute may be significant in a diagnosis.

during the course of a diagnosis. The manner in which the various types of information are employed during a diagnosis will be indicated in the discussion of the inference function and the test selection function below.

2. The Inference Function

The inference function of the program is the means by which the program constructs a current view of the diagnostic problem from the information which constitutes its 'experience' and the attributes which have been detected to date in the study of the patient. The inference function is based on the Bayesian model, and the current view held by the program is a conditional probability distribution for the various diseases. The basic means for obtaining this distribution is discussed in Figure 1.

Whenever new test results become available, the current view of the problem is updated in this manner. It is this updated view which the test selection function employs in selecting new tests to perform. Basically, the operation of the program is as follows. The physician defines a problem for the program by indicating some set of initial attributes. From this set of initial attributes, the program obtains a current distribution (the probability distribution which constitutes the current view of the problem). The program has two alternatives available to it at this point. It can select a test for the physician to perform, or it can make a final diagnosis and cease testing. In order to make this decision, the program invokes the test selection function. In the event that this function selects a test to perform, the program requests that the

user run the test. When the results become available, the inference function is again invoked to update the current distribution. Figure 2 presents a schematic representation of this process. We now turn our attention to the manner in which the test selection function accomplishes its decision-making task.

3. The Test Selection Function

At each stage in a diagnostic problem, the program must decide whether to cease testing and make a terminal decision about the problem or to select a new test for the user to perform. In the event that the latter alternative is chosen, the particular test to be run must be determined. In dealing with this decision problem, the program makes use of several types of information. The first of these is the current distribution which has been developed by the inference function. This current view of the problem will strongly influence expectations about the relative merits of different testing strategies. The second source of information is the probabilities which constitute a major part of the program's medical experience. These probabilities are used in determining the likelihood of various test results. The costs of the relevant tests and the costs of possible misdiagnoses also have a direct bearing on this decision. Finally there is the history of the diagnosis to date, including such information as which tests already have been performed. The model of the test selection function employed by the program provides a rational means for applying all this information to the problem of ranking the various decision alternatives.

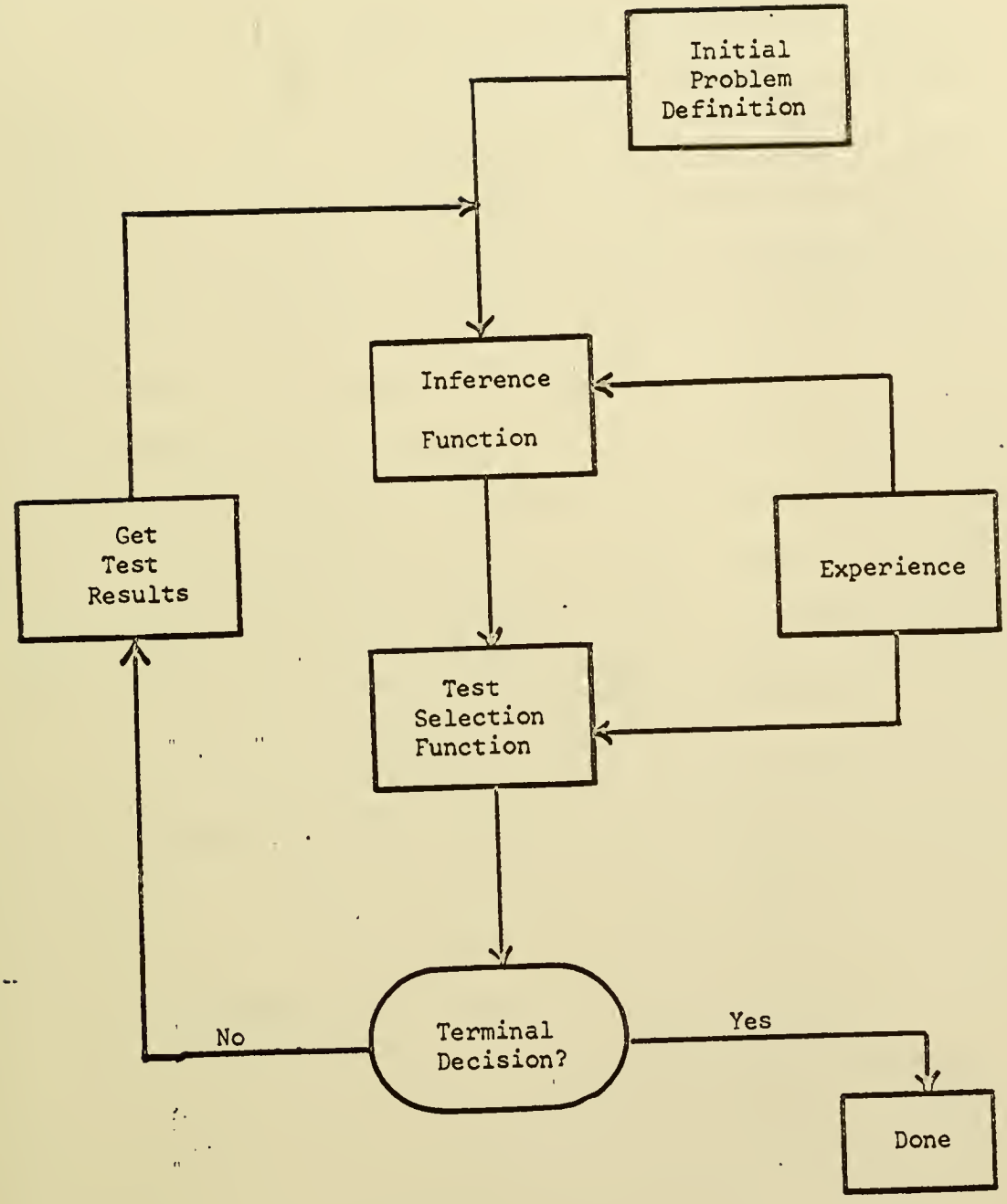

Figure 2

Schematic of Program Flow

Assume that the program is currently confronted with this test selection problem. The position of the program with respect to this problem can be thought of in terms of the decision tree shown in Figure 4. In this abstraction of the decision problem, the program is currently at the node denoted by 1 in the tree. This node is called a decision node, because it is here that the program must make a decision. This decision node is characterized by the current distribution shown in Table 4. There is a branch emanating from this node for each of the decision alternatives available to the program at this point in the diagnosis. Thus there is a branch corresponding to the selection of each of the relevant tests and a branch corresponding to the termination of the diagnosis. This last alternative will be referred to as the terminal decision for the node.

Imagine that a certain test has been selected. This test will yield one of a number of results. Notice that a particular test result may correspond to the observation of more than one attribute. In the tree, there is a node corresponding to the selection of this test. Such nodes are called nature's nodes, because the selection of the branch which will be taken away from such a node is not under the control of the decision-maker, but rather depends on a state of nature. There is a branch emanating from this node for each of the possible test results. When the results of the test become known, they will be used by the inference function to update the current view of the problem, and a new decision node with its associated distribution will be reached. At this point, the program will again face the test selection problem. Because of the many alternative testing strategies encountered in a typical diagnostic problem, the full

decision tree is extremely large, containing many decision nodes and branches. All the information required by the program to build and analyze such a decision tree, however, is available in the model.

The current distribution and the conditional probabilities described in Part 1 of this section can be used to obtain the probability of each of the possible results of a given test in the following way: Given a particular test result, the new current distribution can be computed as shown in Figure 1. Notice that the computation of this distribution is exactly the task of the inference function. Thus the inference function is used not only for the incorporation of new observations into the current view of the problem, but also for the prediction of the results of test sequences being considered. Once the distribution for the new decision node has been obtained, the analysis can be repeated. So the probability information specified in the model is sufficient to permit the assessment of probabilities for the possible outcomes of various testing sequences. There remains, however, the problem of employing these probabilities to rank the decision alternatives.

If it were possible to lay out the entire decision tree for the problem, the solution to the test selection problem could be obtained by 'folding back' the tree in terms of expected value. In a typical problem, however, the determination of an optimal testing strategy will be computationally infeasible because of the large number of decision nodes involved. On the other hand, the program can perform a partial analysis of the tree

by restricting the analysis to only a part of the decision tree. There are many ways to limit the growth of the decision tree, and a number of them have been discussed in detail elsewhere. (7) Only one of these ways will be discussed here.

One way to reduce the amount of computation required to analyse the test selection problem is to restrict the depth of the analysis of the decision tree. This corresponds to considering testing sequences of limited length. For example, only sequences of two tests might be considered. This, of course, introduces the possibility that the program may overlook a very effective sequence of three tests, none of which is particularly useful alone. Whether such a risk is significant can be determined only through the study of the problem area in question. A small example of an analysis of a decision tree to a depth of one is presented in Figure 3. The steps in this analysis are:

1. Obtain the expected cost of a terminal decision for the current decision node. This expected cost is computed using the costs of possible misdiagnoses and the current distribution. By computing the expected cost of misdiagnosis for each of the possible terminal decisions (diseases), the best terminal decision can be determined.

2. For each relevant test, perform the following analysis. First determine the possible results of the test and for each result the probability that the result will be

obtained. These probabilities are computed in accordance with equation 4 of Figure 3. For each test result, determine the new current distribution conditional upon the observation of that result. As was noted above, this distribution can be developed by the inference function through the application of Bayesian analysis.[2] As in Step 1, determine the expected cost of the best terminal decision at this new decision node.

3. Determine the expected cost of a terminal decision following the application of the test in question as follows. Sum the products of the probability of a given test result and the expected cost of the best terminal decision for the decision node corresponding to that result. This sum will be called the expected decision loss for the test. Its use is based on the assumption that the desirability of attaining a particular node in the tree (i.e., a particular view of the problem) can be approximated by the expected cost of a terminal decision at that node. A low value of this cost at a given node means that little risk is incurred in making the diagnosis in accordance with

---

[2]This simply involves simulating the observation of the attribute or attributes in question.

Figure 3

Sample Test Selection Problem

The information contained in Table 1, 2, and 3 will be used in this example of the decision process employed in the test selection function. Assume that the current distribution is

$$P' = P(D1, D2, D3) = (0.4, 0.1, 0.5)$$

The analysis of the decision tree is restricted to a depth of one.

1. Obtain the expected cost of the best terminal decision for the node, EDC,

$$(3) \qquad EDC = \min_{i=1,3} \{EDC_i = \sum_{j=1}^{3} L_{i,j} \cdot P'_j\}$$

where $L_{i,j}$ is an element from the decision loss matrix of Table 3. We find

$$EDC_1 = 350, \; EDC_2 = 450, \; EDC_3 = 270$$

Therefore the best terminal decision at this node is D3 with expected cost 270.

2. Now each of the relevant tests is analyzed as follows. The probability of each test result is computed.

$$(4) \qquad P(S_k) = \sum_{j=1}^{3} P'_j \cdot P_{k,j}$$

where $P_{k,j}$ is a conditional probability from the matrix of Table 1. The relevant tests, their costs, and their possible results are shown in Figure 4. The probability of each test result is indicated on the branch of the tree corresponding to that result. The current distribution for each of the new decision nodes is computed through the use of the inference function. These distributions and the terminal decision and its expected cost for each decision node are listed in Table 4.

Figure 3 (Continued)

3. Let C_l be the cost of the lth test, and ETC be the total expected cost of performing a test and then making an optimal decision based on the results obtained. Then

$$(5) \qquad ETC_l = C_l + \sum_k P\{S_k\} \cdot Q_k$$

where S_k is a possible result of test T_l and Q_k is the expected cost of an optimal decision for the decision node reached on the branch corresponding S_k.

Then the decision alternative selected for the current node is determined from:

$$\min_{l=0,4} \{ETC_l\}, \text{ where } ETC_0 = EDC$$

For our example, we find that $ETC_0 = EDC = 270$

$$ETC_1 = 1. + 0.048 \cdot 238 + 0.952 \cdot 274 = 273.2$$

$$ETC_2 = 3. + 0.565 \cdot 166 + 0.435 \cdot 348 = 248.2$$

$$ETC_3 = 6. + 0.180 \cdot 83 + 0.820 \cdot 312 = 277.0$$

Therefore, the decision chosen by the test selection is to perform the test T2.

Table 3

Decision Losses

	Actual Disease		
Diagnosis	D1	D2	D3
D1	0	1000	500
D2	500	0	500
D3	300	1500	0

Table 1

Conditional Probabilities

	Attributes*			
Disease	S1	S2	S3	S4
D1	0.02	0.45	0.10	0.90
D2	0.05	0.10	0.90	0.10
D3	0.07	0.75	0.10	0.90

*Attributes S3 and S4 are mutually exclusive.

Table 2

Relevant Tests

Test	Attributes	Cost
T1	S1	1.0
T2	S2	3.0
T3	S3, S4	6.0

the associated current distribution. Similarly, a high expected cost for a terminal decision implies high risk. Therefore the program seeks to attain decision nodes which exhibit low expected decision losses. Other measures can be used, although this measure possesses all the desirable properties. Alternative methods for assessing the 'value' of decision nodes are discussed elsewhere. (7)

4. Once the expected decision loss has been determined for each of the relevant tests, the test selection function can make its decision. If a particular test is performed, the cost of the test is incurred. Now that the expected decision loss given the test results is known, the total expected loss for the test can be taken as the sum of these two costs. Recall that the model assumes that test costs and decision losses are additive. The total expected loss for the terminal decision at the current node is expected cost of the best terminal decision at the node. The test selection function simply selects the decision alternative with the minimum total expected loss.

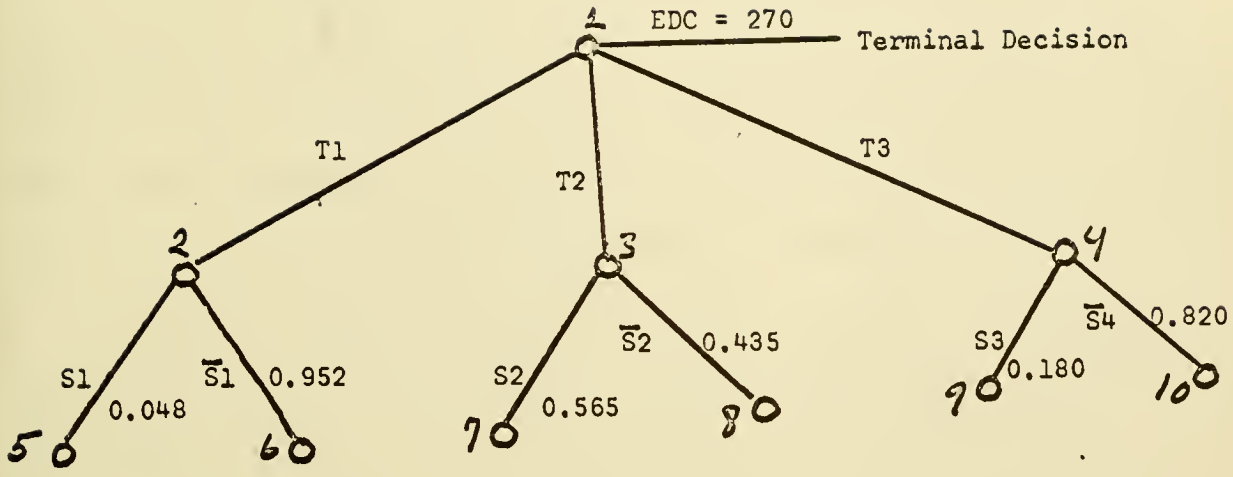

Figure 4

Decision Tree for Sample Problem

Table 4

Results of Analysis

Node	Distribution	Best Decision	EDC
1	(0.4, 0.1, 0.5)	D3	270
5	(0.166, 0.105, 0.729)	D3	238
6	(0.412, 0.100, 0.488)	D3	274
7	(0.318, 0.018, 0.664)	D3	166
8	(0.506, 0.201, 0.293)	D2	348
9	(0.031, 0.834, 0.135)	D2	83
10	(0.429, 0.122, 0.449)	D3	312

Using this scheme, the test selection function will continue to select tests to be performed as long as at least one test is expected to reduce the decision loss by an amount exceeding the cost of the test. Only when no test is expected to reduce the decision loss sufficiently to justify its expense, does the function make a terminal decision.

Although a specific realization of the test selection function has been discussed here, the underlying model will support many other such realizations. Different methods for developing the tree and assigning value to decision nodes can be employed without altering the basic method of the analysis.

C. An Experiment in Sequential Diagnosis

A program based on the model of sequential diagnosis has been implemented as part of a larger system for the study of computer-aided diagnosis. The program operates on the time-sharing system at Project MAC. This system services many users at once, each user communicating with it via a remote terminal. With such a system, it is possible to employ highly interactive programs which engage in a continual dialogue with the user. The diagnostic program was designed to exploit this interactive capability of the time-sharing system.

The performance of the diagnostic program has been studied in two problem areas, the diagnosis of bone tumors and the diagnosis of congenital heart disease. The program was very successful in the limited number of

bone tumor cases presented to it. (7) Because its performance was less successful in the original limited set of heart disease cases considered, it was felt that this latter area better justified an extensive study.

Dr. Homer Warner and his associates generously supplied the probability matrix and a priori disease probabilities required by the program. The matrix reflects the extensive statistical data collected by these workers modified somewhat in accordance with their experience with these diseases. In addition to this data, these investigators also provided a set of several hundred case histories for use in the study.

Because discussions of this particular diagnostic problem are available in the literature (2, 3), only a brief outline of the problem will be presented here. The probability matrix consists of 35 congenital heart disease entities (including 'normal') and 53 attributes. These attributes include 25 heart murmers, 11 X-ray findings, 7 EKG findings, 6 findings from a phonocardiographic tracing, and 3 age groups. No data from dye-injection studies or heart catheterization were employed.

The case histories included the attributes which were present and the definitive diagnosis. In order to permit the program to exploit the sequential approach to diagnosis, thirty-four tests were defined. These tests are employed by the program to elicit information about a case. For example, one of the tests is used to discover the age group of the patient; another to ascertain the presence and extent of cyanosis; and a third, to detect cardiomegaly.

An experiment in sequential diagnosis was performed under the following conditions. The cost of each test was taken to be 1.0, and the cost of each possible misdiagnosis was set to 1000. For each case, the program was given the results of the same three tests as the initial definition of the problem. These test results revealed the age group of the patient, whether he was cyanotic, and whether the squatting attribute was present. Once the program had received the initial definition of the problem, it entered the inference-test selection loop depicted in Figure 2. The decision tree analysis employed in the test selection function was exactly that discussed in Figure 3. The current distribution at the time of a terminal decision was recorded as well as the testing strategy employed to reach this diagnosis. Then the results of the tests not selected by the program were revealed to the inference function. The resulting distribution, called the complete diagnosis, therefore, is based on all the available signs for the case. The results of this experiment are summarized in Table 6 where

$\bar{P}_s$, $\bar{P}_c$ are the average probabilities assigned to the actual disease in the sequential diagnosis and in the complete diagnosis respectively.

f_s, f_c are the fractions of the cases in which the actual disease was assigned a probability greater than 0.01 in the respective diagnoses.

$\bar{n}$ is the average number of tests selected (in addition to those used in the original problem definition) by the program in sequential diagnosis.

Warner, et al., suggest in (3) that the product ($Z = \bar{P} \times f$) can be taken as a measure of diagnostic performance. This measure reflects both the average probability assigned to the actual disease and the relative number of complete failures. For convenience, this measure is used here, and values of it are also included in the table.

An expected performance measure, $\bar{Z}$, can be obtained by taking a weighted average of the individual performance measures

$$\bar{Z} = \frac{\sum_{\text{all } j} P'_j Z_j}{\sum_{\text{all } j} P'_j}$$

Table 6

Summary of Experimental Results

Disease*	P'	Cases	$\bar{P}_s$	f_s	Z_s	$\bar{n}$	$\bar{P}_c$	f_c	Z_c
1	0.1588	43	0.79	1.00	0.79	9.5	0.84	1.0	0.84
2	0.1597	37	0.74	0.95	0.72	5.5	0.83	0.90	0.74
3	0.0316	2	<.01	0.00	0.00	8.5	<.01	0.00	0.00
4	0.0183	4	<.01	0.00	0.00	6.0	<.01	0.00	0.00
5	0.0849	9	0.93	1.00	0.93	1.9	0.96	1.00	0.96
7	0.0133	2	0.12	1.00	0.12	15.5	0.03	0.50	0.02
8	0.0017	1	0.02	1.00	0.02	7.0	0.93	1.00	0.93
10	0.0206	6	0.07	1.00	0.07	5.0	0.04	0.67	0.03
11	0.0475	11	0.64	0.91	0.58	5.9	0.64	0.91	0.58
12	0.0392	3	0.64	0.67	0.40	9.0	0.63	0.67	0.42
13	0.0316	2	<.01	0.00	0.00	10.0	<.01	0.00	0.00
15	0.0017	2	<.01	0.00	0.00	4.0	<.01	0.00	0.00
16	0.0017	3	0.09	0.33	0.03	10.7	0.14	0.67	0.09
18	0.0549	16	0.58	0.88	0.51	6.3	0.44	0.63	0.28
23	0.0366	24	0.94	1.00	0.94	5.0	0.99	1.00	0.99
24	0.0750	17	0.88	0.93	0.81	6.0	0.90	1.00	0.90
26	0.0033	4	<.01	0.00	0.00	6.5	<.01	0.00	0.00
27	0.0050	2	0.46	0.50	0.23	6.5	0.44	0.50	0.22
30	0.0548	27	0.64	0.80	0.51	5.8	0.66	0.72	0.48
31	0.0033	8	0.46	0.63	0.29	9.1	0.44	0.75	0.33
32	0.0150	5	0.004	0.25	0.001	9.6	0.03	0.25	0.008
34	0.0606	15	0.14	0.87	0.12	11.8	0.23	0.87	0.20
	0.9271	245							

*The diseases are numbered in accordance with (3). Notice that there were no cases of certain diseases studied.

where P'_j is the a priori probability of disease 'j' and j indexes all diseases for which at least one case was studied. For the 245 cases studied,

$$\sum_{\text{all } j} P'_j = 0.93$$

Similarly, the expected average number of tests for sequential diagnosis, $\bar{n}$, can be obtained from the average number of tests for each disease, $\bar{n}_j$ by:

$$\bar{n} = \frac{\sum P'_j \bar{n}_j}{\sum P'_j}$$

For this experiment, these values are: $\bar{Z}_s = 0.57$, $\bar{Z}_c = 0.59$, $\bar{n} = 6.9$.

D. Discussion of Results

Although the results presented above are somewhat limited, they do provide some interesting insights into the potential usefulness and limitations of this model of sequential diagnosis. Perhaps the most striking aspect of these results is the comparison of sequential and complete diagnoses which they provide.

As indicated by the performance measures $\bar{Z}_s$ and $\bar{Z}_c$, there is essentially no difference in the expected average accuracy of the two methods. The complete diagnosis employed in these tests corresponds to the use of the Bayesian model in the program developed by Warner. Because Warner has determined that this latter use of the model obtains results comparable to

those experts (3), it is expected that the sequential program would also attain this expert level. Although the complete diagnosis and the sequential diagnoses provide the same expected accuracy as measured by $\bar{Z}$, the former employs all the available test results, while the latter employs only 6.9 tests on the average. With the definition of tests used in this experiment, thirty-one tests are required to obtain all available results. Thus there is a sharp reduction in the average number of tasks required in sequential diagnosis. This reduction results from the fact that the test selection function chooses tests based on a continually updated view of the problem. Therefore only those tests which are expected to improve the position of the program with respect to the problem are selected. Other tests are ignored. When no test is expected to provide a reduction in expected decision loss in excess of its cost, the terminal decision is made. Because the two measures $\bar{Z}_s$ and $\bar{Z}_c$ are nearly equal, it appears that the test selection function is generally correct in its assessment of the value of further testing. It is interesting to note that this success was achieved even though the test selection function restricted the depth of the decision tree analysis to one.

The heart disease diagnosis problem was chosen to test the program because the required statistical data were available and because it is a non-trivial problem. It did provide, however, a good example of the potential value of the sequential model. Because the costs of medical tests (in terms of patient discomfort, time of skilled persons, money, etc.) may

be significant, the use of inefficient testing sequences should be regarded as ineffective diagnoses. This experiment provides an example of the value of a few, well-chosen tests from the set of available tests in the diagnosis of heart disease.

In attempting to evaluate this reduction in the average number of tests required for a diagnosis through the use of sequential test evaluation, one is confronted with the problem of the costs employed by the model. The costs for tests and for misdiagnoses used for this experiment were arbitrary. No attempt was made to approximate the actual costs for the problem area. These costs play an important role in the decision-making process of the test selection function. If, for example, the average test cost is of the same order of magnitude as that of a misdiagnosis, the average number of tests selected by the program will decrease. On the other hand, if the average cost of a misdiagnosis greatly exceeds that of an average test, this number will be relatively greater. This, of course, is a desirable property for the model to exhibit. For a given problem area, however, uncertainty about these costs will result in an uncertainty about the exact value of sequential diagnosis. In the heart disease problem, for example, additional study would be required to establish the value of the reduction in the average testing cost obtained from the use of the sequential model.

The costs for possible misdiagnoses are fundamental to a proper evaluation of the performance of the program. While the overall average measures $\overline{Z}_c$ and $\overline{Z}_s$ are approximately equal, there are differences in the values for certain diseases. For example, the pair (Z_s, Z_c) is (0.58, 0.44)

for disease 18 and (0.14, 0.23) for disease 34. If the misdiagnosis of either of these diseases is very serious relative to other possible misdiagnoses, these differences may significantly affect the evaluation of the comparison of sequential and complete diagnoses. Unfortunately, the appropriate cost of misdiagnoses is not known at present, and so the performance measures such as those presented above must suffice.

In spite of these difficulties, the sequential model may prove quite useful in computer-aided diagnosis. This seems particularly true for problems in which the identification of diseases requires very different test sequences. In such a case, many tests may be irrelevant for a given patient. Here, one clearly wants to limit the diagnostic tests as closely as possible to the set of relevant tests. Sequential testing provides a means for accomplishing this.

The sequential model, then, seems to offer a significant advantage for use in computer-aided diagnosis, and further research on its use is planned. A number of interesting aspects of the model require further investigation.

First, as with all Bayesian models, the performance of this model depends on the probabilities it employs. The more accurately these probabilities reflect the statistical properties of the diseases in question, the more accurate the inference of the program will be. Whether this performance is adversely affected by small errors in the probability estimates is a question open to further investigation. Even if the probabilities were known exactly, there would still undoubtedly remain a

for disease 18 and (0.14, 0.23) for disease 34. If the misdiagnosis of either of these diseases is very serious relative to other possible misdiagnoses, these differences may significantly affect the evaluation of the comparison of sequential and complete diagnoses. Unfortunately, the appropriate cost of misdiagnoses is not known at present, and so the performance measures such as those presented above must suffice.

In spite of these difficulties, the sequential model may prove quite useful in computer-aided diagnosis. This seems particularly true for problems in which the identification of diseases requires very different test sequences. In such a case, many tests may be irrelevant for a given patient. Here, one clearly wants to limit the diagnostic tests as closely as possible to the set of relevant tests. Sequential testing provides a means for accomplishing this.

The sequential model, then, seems to offer a significant advantage for use in computer-aided diagnosis, and further research on its use is planned. A number of interesting aspects of the model require further investigation.

First, as with all Bayesian models, the performance of this model depends on the probabilities it employs. The more accurately these probabilities reflect the statistical properties of the diseases in question, the more accurate the inference of the program will be. Whether this performance is adversely affected by small errors in the probability estimates is a question open to further investigation. Even if the probabilities were known exactly, there would still undoubtedly remain a

for disease 18 and (0.14, 0.23) for disease 34. If the misdiagnosis of either of these diseases is very serious relative to other possible misdiagnoses, these differences may significantly affect the evaluation of the comparison of sequential and complete diagnoses. Unfortunately, the appropriate cost of misdiagnoses is not known at present, and so the performance measures such as those presented above must suffice.

In spite of these difficulties, the sequential model may prove quite useful in computer-aided diagnosis. This seems particularly true for problems in which the identification of diseases requires very different test sequences. In such a case, many tests may be irrelevant for a given patient. Here, one clearly wants to limit the diagnostic tests as closely as possible to the set of relevant tests. Sequential testing provides a means for accomplishing this.

The sequential model, then, seems to offer a significant advantage for use in computer-aided diagnosis, and further research on its use is planned. A number of interesting aspects of the model require further investigation.

First, as with all Bayesian models, the performance of this model depends on the probabilities it employs. The more accurately these probabilities reflect the statistical properties of the diseases in question, the more accurate the inference of the program will be. Whether this performance is adversely affected by small errors in the probability estimates is a question open to further investigation. Even if the probabilities were known exactly, there would still undoubtedly remain a

certain probability of error in most problem areas. Thus there is an upper bound on the performance of such a model. How serious this limitation is in a particular area is again an open question. Mount and Evans (8) have performed studies which are relevant to this problem.

The problem of costs for the model has been raised before, but it bears repeating that the performance of the model can best be measured in terms of these costs. The fact that these costs directly affect decisions made by the test selection function provides additional motivation for establishing them. Of course it will be extremely difficult to determine costs for possible misdiagnoses and various tests, all on a common scale. One approach is to test the sensitivity of the model to these costs in order to isolate those which most affect performance in a given area. Then attention can be focused on these.

Despite the difficult problems which remain, it is believed that the basic model of diagnosis presented here merits considerable study as a reasonable means for exploring computer technology in a variety of diagnostic problems. Naturally, the model will require modifications as the problems of computer-aided diagnosis become better understood, but it is believed that this model can accommodate the necessary changes.

REFERENCES

(1) Ledley, R. S., and L. B. Lusted, "Reasoning Foundation of Medical Diagnosis: Symbolic Logic, Probability, and Value Theory Aid Our Understanding of How Physicians Reason", Science 130, 9, 1959.

(2) Toronto, A. F., L. G. Veasy, and H. R. Warner, "Evaluation of a Computer Program for Diagnosis of Congenital Heart Disease", Progress in Cardiovascular Diseases, 5, 362, 1963.

(3) Warner, H. R., A. F. Toronto, and L. G. Veasy, "Experience with Baye's Theorem for Computer Diagnosis of Congenital Heart Disease", Annals of New York Academy of Sciences 115, 2, 1964.

(4) Overall, J. E., and C. M. Williams, "Conditional Probability Program for Diagnosis of Thyroid Function", Journal of the American Medical Association, 183, 5.

(5) Lodwick, G. S., et al., "Computer Diagnosis of Primary Bone Tumors: A Preliminary Report", Radiology, 80, 1963.

(6) Gorry, G. A., "Strategies for Computer-Aided Diagnosis", Mathematical Biosciences (in press).

(7) _____, "A System for Computer-Aided Diagnosis", Project MAC Technical Report, M.I.T. (in press).

(8) Mount, J. F. and J. W. Evans, "Computer-Aided Diagnosis -- A Simulation Study", Fifth IBM Medical Symposium, 1963.

Date Due